



HAL
open science

Estimating the prevalence of child labour in the cocoa industry via indirect elicitation methods: a mixed-methods study

Aurelia Lepine, Ariane Ndiore, Carole Treibich, Henry Cust, Laurent Foubert, Megan Passey, Selina Binder

► To cite this version:

Aurelia Lepine, Ariane Ndiore, Carole Treibich, Henry Cust, Laurent Foubert, et al.. Estimating the prevalence of child labour in the cocoa industry via indirect elicitation methods: a mixed-methods study. *Journal of Population Economics*, 2024, 37 (4), pp.77. 10.1007/s00148-024-01054-3. hal-04791105

HAL Id: hal-04791105

<https://hal.science/hal-04791105v1>

Submitted on 19 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Estimating the prevalence of child labour in the cocoa industry via indirect elicitation methods: a mixed-methods study

Aurelia Lepine¹ · Ariane Ndiore² · Carole Treibich³ · Henry Cust¹ · Laurent Foubert⁴ · Megan Passey⁴ · Selina Binder⁵

Received: 18 May 2023 / Accepted: 14 October 2024
© The Author(s) 2024

Abstract

Data from children suggest that the prevalence of child labour in the cocoa industry in Côte d'Ivoire is a worrying 38%. However, while surveying children has the potential to reduce sensitivity biases such as social desirability bias, it presents significant ethical dilemmas and may also be associated with other reporting biases, making accurate estimates of child labour difficult. To address this, we collected data from 1741 cocoa producers to assess whether parents could provide more accurate estimates using indirect survey methods. We compared direct questioning with a list experiment and a novel non-verbal method ('colorbox'). We found higher prevalence rates of child labour using indirect elicitation methods, but lower than those obtained from children's surveys. Qualitative in-depth interviews revealed motivations for underreporting, including fear of legal consequences and mistrust of stakeholders. Indirect methods alone are not sufficient to correct for underreporting when child labour is collected from parents. Future research should prioritise direct data collection from children and address ethical concerns to obtain more accurate estimates of child labour.

Keywords Child labour · Sensitive bias · Measurement error · Cocoa · Sustainable development goals

JEL Classification J80 · J13 · J22 · O15 · I25

Responsible editor: Kompal Sinha

Extended author information available on the last page of the article

1 Introduction

According to the International Labour Organization, the term ‘child labour’ is often defined as ‘work that deprives children of their childhood, their potential and their dignity, and that is harmful to their physical and mental development’. While child labour can take many forms, a priority for the international community is the elimination of the worst forms of child labour, as defined in Article 3 of ILO Convention No. 182, which includes hazardous work, i.e., work that is harmful to the health and safety of the child. In 2010, the government of Côte d’Ivoire, representatives of the cocoa industry and the US Department of Labor signed a declaration in which the industry publicly acknowledged the existence of child labour in the cocoa sector and pledged to address it and achieve a 70 per cent reduction in child labour in the cocoa sector by 2020. Despite many interventions by government, industry and international organisations, tracking this progress is hampered by serious measurement error in estimating the prevalence of child labour.

In such a context, measurement bias can be caused by several factors. Côte d’Ivoire is the world’s largest cocoa producer, supplying around 40% of the world’s cocoa beans (Our World in Data., 2021). Côte d’Ivoire has enacted laws and regulations prohibiting child labour, including setting a minimum working age and defining a list of hazardous activities in line with international conventions (Côte d’Ivoire United States Department of Labor (DOL), 2015). Significant efforts have also been made to raise awareness of the dangers of child labour among the rural population (Government of Côte d’Ivoire et al. 2021). Farmers can be certified for meeting ethical and sustainability standards demanded by consumers, including the absence of child labour. Certification leads to a legal price and certification premium that rewards those who do not use child labour. This knowledge, combined with the fear of criminal sanctions or withdrawal of certification, may encourage under-reporting of child labour by cocoa farming households. A second problem with monitoring child labour through data collection is social desirability bias, where farmers are reluctant to disclose their behaviour to enumerators in face-to-face interviews. Thirdly, another type of sensitivity bias is caused by parents who may have a negative self-image if their children are working and therefore practice self-denial (Nederhof 1985). The different sources of sensitivity bias are presented in Appendix 1.

In this paper, we collected data to elicit three indicators of child labour from parents of children aged 5–17 living in agricultural households producing cocoa, using two indirect elicitation methods. We estimate whether and to what extent child labour is underreported. We also examine the sources of potential reporting bias and the extent to which the indirect elicitation methods account for this bias through qualitative interviews. Finally, we estimate the determinants of child labour using our indirect elicitation methods compared to direct interviews, in order to examine quantitatively what might cause bias in child labour reporting.

We expect that parental underreporting may be due to a number of factors, including the financial implications of losing certification, legal sanctions (such

as fines and imprisonment), threats to personal safety, social desirability bias, and negative self-image. It is important to note that the effectiveness of available approaches to collecting information on child labour may vary depending on the sources of sensitivity bias. For instance, non-verbal methods, methods that do not allow identification of individuals engaging in the sensitive behaviour, or data collection from children may be more or less effective in overcoming sensitivity bias. Methods that introduce random noise into the response, such as list experiments, may allow us to overcome most sources of bias. Non-verbal methods (e.g., colorbox) will primarily allow correction of social desirability bias if responses can be linked to respondents and will be as effective as the list experiment if data are collected anonymously. However, in the presence of self-denial due to negative self-image, indirect elicitation methods may be less effective in correcting for sensitivity bias. It is therefore important to identify the sources of sensitivity bias conceptually in order to select the most appropriate indirect elicitation methods and to use qualitative methods to assess respondents' perceptions of the confidentiality of the chosen indirect elicitation method. Typically, indirect methods are used without explaining to participants how they work, and perceptions of confidentiality may often not reflect the true confidentiality of the method.

The sample was drawn from four major cocoa-growing regions in Côte d'Ivoire. The measures include indicators of hazardous child labour, which the International Labour Organization (ILO) considers to be one of the worst forms of child labour. Our first indirect survey method is the list experiment or item count method. In this method, participants are asked how many statements they agree with, without revealing which ones. Participants are randomly assigned to receive treatment or control lists. The control list of statements contains three non-sensitive statements, while the treatment list contains the same three non-sensitive statements plus one sensitive statement. In our case, this was about the use of child labour in cocoa production. By comparing the difference in the average number of statements agreed with between the treatment and control lists, we can determine the average number of people who agreed with the sensitive statement, and therefore the prevalence of that statement across the whole sample. We conducted the more robust double-list experiment, which repeats the same method except that the treatment and control allocations and the non-sensitive statements are reversed. This allows the results of the two to be pooled, minimising standard errors and providing more precise estimates (Lépine et al. 2020). The second method used was a non-verbal method developed by our research team called 'colorbox'. In this method, participants were given coupons with a colour associated with their answer (white for 'yes' and black for 'no') and an associated 4-digit code. Participants had to select the coupon with the colour corresponding to their answer and give the PIN code to the enumerator. This method avoided direct disclosure of the response to the enumerator, while still allowing the correct response to be associated with the participant. The above methods were selected and the design of the quantitative survey was developed following the completion of 36 in-depth interviews with farmers. The topics covered in the interview guide were designed to (i) assess the willingness of farmers to disclose instances of child labour and the underlying causes to different stakeholders and (ii) assess the perception of confidentiality of the two indirect survey methods.

The results of our study suggest that child labour is under-reported by cocoa farmers. The list experiment and colorbox methods yielded about twice the prevalence of child labour questions compared to asking farmers directly. Qualitative evidence suggests a high level of knowledge about the harms of child labour and related sanctions. However, farmers expressed distrust of many external actors, highlighting the importance of using our indirect survey method. Furthermore, our findings suggest that child labour prevention interventions may inadvertently discourage the reporting of child labour, possibly due to increased awareness of associated sanctions, rather than actually reducing prevalence. This is an important consideration, as misreporting bias has been shown to influence conclusions about the effectiveness of child labour reduction interventions (Lichand and Wolf 2022).

Apart from Jouvin (2023), there is a paucity of literature on the collection of child labour statistics with sensitivity bias taken into account. Data collected from different stakeholders is obtained through direct questioning, which is susceptible to this sensitivity bias. Consequently, figures collected from stakeholders in the cocoa industry (e.g., local governments and cooperatives) are more likely to suffer from sensitivity bias and should not be used to track progress in reducing child labour due to these measurement errors. Furthermore, the literature highlights a number of additional considerations in the collection of accurate child labour statistics. These include differences in survey instruments (Dillon et al. 2012; Guarcello et al. 2010), recall periods (Beegle et al. 2012) and the identity of the respondents (Dillon 2010), abortion (Bell and Bishai 2019; Moseson et al. 2021), voting preferences (Gonzalez-Ocantos et al. 2012; Holbrook and Krosnick 2010), use of microfinance loans (Karlan and Zinman 2012), opinions on undocumented immigrants (McKenzie and Siegel 2013), gay marriage (Lax et al. 2016), racism (Krumpal 2013), intimate partner violence (Agüero and Frisancho 2022; Asadullah et al. 2021; Castilla and Murphy 2023; Cullen 2023; Gilligan et al. 2024) and condom use (LaBrie and Earleywine 2000; Treibich and Lépine 2019). The novel colorbox method has not yet been widely adopted, but has been used to collect data on risky sexual behaviour among sex workers (Lepine et al., 2024) and on fertility preferences (Valente et al. 2024).

This paper contributes to several strands of literature, primarily by adding to the evidence on child labour in cocoa production in Côte d'Ivoire (Jouvin 2023; Sadhu et al. 2020) and the most accurate methods for collecting such information. This study contributes to the emerging literature on the reliability of the double-list experiment method for obtaining accurate prevalence estimates (Haber et al. 2018; Bell and Bishai 2019; Chuang et al. 2021). It also contributes to the literature on the feasibility and validity of the colorbox method using the list experiment as a benchmark (Lépine et al. 2024). Methodologically, our approach builds upon that of Jouvin (2023) by utilising the double-list experimental design. This design allowed us to assess the internal consistency of the child labour prevalence estimates derived from the two lists, to confirm the validity of our list experiment design, and to investigate in more detail the determinants of sensitivity bias. The paper underscores the critical importance of accurately measuring child labour in the context of the Sustainable Development Goals (SDGs). It emphasizes how robust measurement is essential for tracking

progress towards SDG 8.7 (elimination of child labour and forced labour). By highlighting the significance of measurement, the paper contributes to the global discourse on child labour within the framework of sustainable development.

2 Methods

We focused on measuring children aged 5–17 working in cocoa-growing areas in four regions of Côte d'Ivoire where cocoa farming is common: Bianouan and Afféry (south-east), Sinfra (centre) and Bloléquin (west) (see Fig. 1). We chose these areas because they are major cocoa-growing regions (Chingandu et al. 2017) and reflect geographical differences as well as differences in organisational development of cooperatives and child labour interventions. Data were collected between January and March 2022, at the end of the main cocoa campaign (September to December) and before the small campaign (February to April), when farmers were more available. 'Campaign' is a direct translation of the terms used to describe the growing seasons or cropping cycle in Côte d'Ivoire, but the term differs from 'campaigns' used in reference to information or education programmes. The project was authorised by the ARTCI, the official public authority responsible for authorising the collection and processing of personal data.

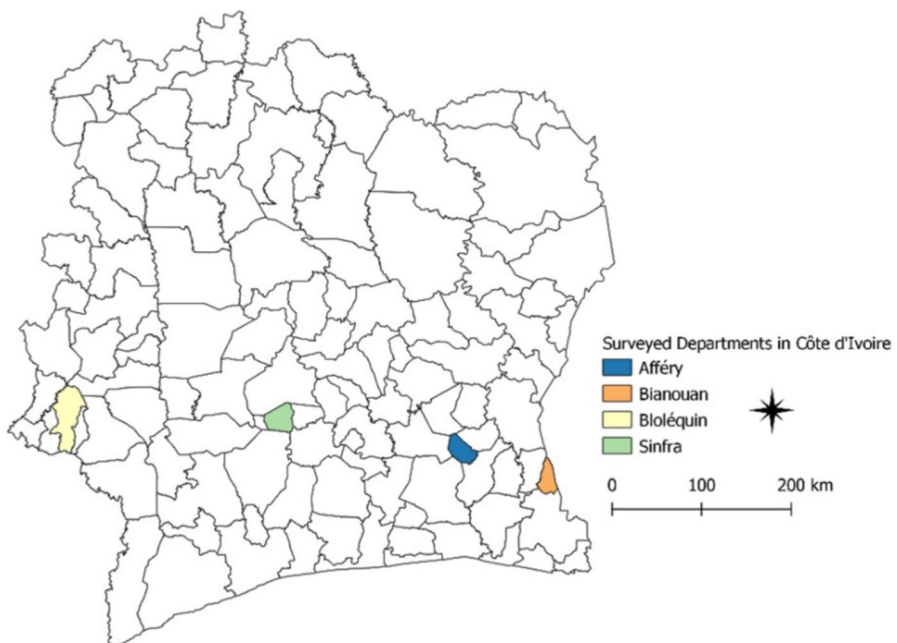


Fig. 1 Surveyed departments in Côte d'Ivoire

2.1 Qualitative study

Data were collected in villages and camps with cocoa farmers who had at least one child aged 5–17 in their household. Farmers were pre-selected from cooperative lists (which included farmers who were not members of a cooperative) within the four selected regions. This ensured a proportional representation of female farmers (around 20%) and a spread of wealth and time in and out of cooperatives. This resulted in 36 face-to-face interviews being conducted in January and February 2022. The interviews were recorded with the consent of the participants using voice recorders. Participants were pre-selected using a census of farmers from the main cooperatives in the area. Several characteristics were balanced, including gender and participation in awareness-raising sessions on child labour and certification. Once in the village, the lead anthropologist researcher met with cooperative representatives to identify the pre-selected farmers and facilitate the interviews. Interviews were conducted in the local language and facilitated by an independent translator using a semi-structured interview guide developed by our research team.

The interviews began with open-ended questions about attitudes to child labour, covering two main topics:

- 1) Questions about willingness to disclose child labour in general and to specific stakeholders (cocoa company, cooperative, peer farmers, government, police, enumerator and researchers),
- 2) Questions about the risks and benefits to themselves and to children working on cocoa farms.

Finally, we also piloted the list experiment and colorbox methods and gathered perceptions about the confidentiality of the methods. The interviews lasted 45 to 60 min each, and participants were paid CFAF 3000 (~US\$5) for their time.

The recordings of the in-depth interviews were then translated and transcribed into French by the authors, and thematic analysis was carried out by coding the data for relevant themes and verbatim quotations. This analysis was then triangulated with the quantitative findings to provide a more complete and nuanced understanding of the quantitative findings.

2.2 Quantitative study

We surveyed 1741 farmers located in the four selected areas in March 2022. Villages within the four selected regions were visited, and farmers with children aged 5–17 years old were selected using a random walk through the area covering the majority of the farmers within the village. The sample is not designed to be representative of the country but largely reflects the composition of the majority of farmers producing cocoa in Côte d'Ivoire.

The interviews were conducted by female professional enumerators recruited by the research team, as farmers in the qualitative research indicated that they were

more willing to discuss child labour with women, as they were seen as the ones responsible for the children. The enumerators were from Abidjan but were selected for their knowledge of local languages spoken in the area. The cooperatives facilitated recruitment, but the interview was conducted on the farmer's premises, and it was made clear that our research team was independent. Once eligibility had been established and consent had been obtained from the households, the quantitative survey proceeded to ask about the children in the household and the head of the household. This was followed by indirect survey methods, which were then translated into direct questions about child labour. The head of the household was also asked whether he or she had participated in any awareness campaigns on child labour. The topic of child labour was not explicitly mentioned; instead, the survey was described as focusing on 'challenges of cocoa farming' to avoid priming the topic for the indirect elicitation questions. The average interview lasted 20–30 min, and participants were paid CFAF 2000 (\$3–\$4) for their time.

The paper uses two methods to overcome sensitivity bias: colorbox and the list experiment. Table 1 summarises the strengths and weaknesses of the list experiment and the colorbox methods. The list experiment is used as the benchmark because it provides confidentiality to both the interviewer and the researcher analysing the data. However, the prevalences derived from the list experiment have a wider confidence interval, as the method does not allow information to be retrieved at an individual level. This is the reason why we developed the colorbox method (Lépine et al. 2024). This method allows researchers to collect sensitive information non-verbally and, after anonymising the data, to identify responses at the individual level, thus allowing individual-level analysis. In addition, we also collected actual child labour indicators using an open-ended question to estimate the extent of misreporting due to social desirability bias.

Participants were randomised to the list experiment or to the colorbox method: two-thirds of the sample were asked to elicit child labour using the list experiment method ($n = 1190$), while one-third used colorbox ($n = 551$) with a proportional balance maintained across geographical zones. The decision to oversample farmers in the list experiment group was justified by the imprecision of the list experiment method, which requires a larger number of observations to obtain a prevalence with reasonable confidence intervals. To the best of our knowledge, there is no power calculation that deals with non-individual level data, but we believe that our sample size provides sufficient power as previous list experiments (single and double sided) have been conducted to detect small prevalences with smaller sample sizes than ours (Chuang et al. 2021; Lépine et al. 2020). After implementing the method, we collected participants' perceptions regarding some aspects of the method used (e.g., objective and confidentiality). In both groups, we also collected actual child labour indicators using an open-ended question to estimate the level of misreporting after using the indirect elicitation method.

Our study followed the definition of child labour in cocoa activities provided by the NORC framework, which identifies six categories of hazardous work (conducting land clearing, carrying heavy loads, using agrochemicals, using sharp tools, engaging in long working hours, or working at night). The selection of child labour indicators used in the study was validated by the International Cocoa Initiative, a

Table 1 Strengths and weaknesses of the list experiment and colorbox methods

Method	Description	Strengths	Weaknesses
Colorbox	<p>Respondents use PIN codes associated to a colour that corresponds to their desired answer instead of directly answering to the enumerator</p>	<ul style="list-style-type: none"> - Enumerators do not know with which item respondents agree on - Can be implemented in low literacy settings - Allows analysis at the individual level 	<ul style="list-style-type: none"> - Only accounts for confidentiality issue to the interviewer but responses can be known by researchers - Some logistics involved to implement the method
List experiment	<p>Respondents are allocated randomly to two different groups. They are asked how many of the j non-sensitive items (plus one sensitive item) they agree with if they belong to the control group (the treatment group)</p>	<ul style="list-style-type: none"> - Enumerators and researchers do not know with which items respondents agree on so the method guarantees confidentiality - Can be implemented in low-literacy settings - Allows sub-group analysis 	<ul style="list-style-type: none"> - Does not allow individual-level analysis - Leads to estimated prevalences with larger SE - List experiment variable cannot be a right-hand side variable - Complex design - Some logistics involved to implement the method

not-for-profit foundation that aims to protect the rights of children and adults in cocoa-growing communities in West Africa. Specifically, we defined child labour in the cocoa sector as any child 5–17 years old who works in cocoa farming and is taking part in a hazardous activity. While the NORC framework defines six hazardous activities, we decided to focus on the use of agrochemicals, sharp tools and carrying heavy loads, given that these hazardous activities were the most prevalent in the cocoa sector (Sadhu et al. 2020). These hazardous activities cover almost all child labour cases in cocoa in Cote d'Ivoire. In addition, we included an indicator to capture the child labour of younger children, given its larger negative impact on child development through reduced education and lower physical development to carry out hazardous work (Ibrahim et al. 2019). The three measures of child labour collected in this study were as follows.

1. During this campaign, are there children under the age of 18 who use sharp tools such as machetes or carry heavy loads such as a 10 L water tank on your plots?
2. During this campaign, are there children under 18 who are present when agrochemicals are used on your plots?
3. During this campaign, are there children under 13 who work on your plots?

2.2.1 List experiment

The principle of the list experiment is to randomly assign respondents to two different groups: a control group and a treatment group. Individuals allocated to the control group are presented with a number of non-sensitive statements. They are not asked to say whether they agree with each statement, only how many of the statements they agree with. The same statements are then presented to the treatment group, except that one sensitive statement is added to the set of non-sensitive statements. Assuming that the two groups have similar opinions on the non-sensitive statements, the proportion of individuals in the treatment group who agreed with the sensitive item can be inferred by comparing the average number of statements that respondents in each group agreed with (Blair and Imai 2012; Glynn 2013).

In our surveys, participants in the control (treatment) group were given the following instructions: "I [the interviewer] am going to read three (four) statements. I will then ask you how many of these statements you agree with. You should not tell me which specific statements you agree with, but the number of statements you agree with. I will give you three (four) marbles and you must hold them in your right hand. Keep both hands behind your back. For each statement, if you agree with it, please move one marble from your right hand to your left hand behind you. If you do not agree, please do not transfer a marble. I will not know and please do not inform me. At the end I would like to know the total number of statements you agree with. This number should correspond to the number of marbles you have in your left hand. I will now read out the statements".

We extend this methodology by using two lists instead of one, with each group serving sequentially as the treatment and then the control group, or vice versa (Droitcour et al. 2004; Hadji et al. 2016). More specifically, the same sensitive item was used, but respondents were presented with two different lists of

Table 2 List experimental design

Group 1	Group 2
<p>List 1A:</p> <ol style="list-style-type: none"> 1. I harvested at least one bag of cocoa last campaign 2. During this campaign, children under 18 use sharp tools or carry heavy loads like a water tank of 10 L on my plots 3. The month of December is the toughest month financially for me 4. I harvested more than 18 bags of cocoa last year <p>List 1B:</p> <ol style="list-style-type: none"> 1. I use fertiliser less than 3 times a year 2. The month of January is the toughest month financially for me 3. I have more than 16 years of experience in the cocoa culture <p>List 2A:</p> <ol style="list-style-type: none"> 1. We are a family of cocoa farmers 2. During this campaign, some children under 18 are present on my plots when agrochemical are used 3. My household has more than 30 members 4. At least two members of my household work on my plots <p>List 2B:</p> <ol style="list-style-type: none"> 1. I work on a cocoa plot 2. In addition to cocoa I also grow grapes 3. I have at least five economic dependents <p>List 3A:</p> <ol style="list-style-type: none"> 1. I depend on my work to feed my family 2. During this campaign, there are children under 13 who work on my plots 3. I addition to cocoa I also grow apples 4. I have at least six economic dependents <p>List 3B:</p> <ol style="list-style-type: none"> 1. I am a cocoa farmer 2. I live on my own 3. I carry cocoa beans by motorbike 	<p>List 1A:</p> <ol style="list-style-type: none"> 1. I harvested at least one bag of cocoa last campaign 2. The month of December is the toughest month financially for me 3. I harvested more than 18 bags of cocoa last year <p>List 1B:</p> <ol style="list-style-type: none"> 1. I use fertiliser less than 3 times a year 2. During this campaign, children under 18 use sharp tools or carry heavy loads like a water tank of 10 L on my plots 3. The month of January is the toughest month financially for me 4. I have more than 16 years of experience in the cocoa culture <p>List 2A:</p> <ol style="list-style-type: none"> 1. We are a family of cocoa farmers 2. My household has more than 30 members 3. At least two members of my household work on my plots <p>List 2B:</p> <ol style="list-style-type: none"> 1. I work on a cocoa plot 2. During this campaign, some children under 18 are present on my plots when agrochemical are used 3. In addition to cocoa I also grow grapes 4. I have at least five economic dependents <p>List 3A:</p> <ol style="list-style-type: none"> 1. I depend on my work to feed my family 2. I addition to cocoa I also grow apples 3. I have at least six economic dependents <p>List 3B:</p> <ol style="list-style-type: none"> 1. I am a cocoa farmer 2. During this campaign, there are children under 13 who work on my plots 3. I live on my own 4. I carry cocoa beans by motorbike

non-sensitive items. The order of the list items was identical for all respondents, and everyone received list A first and list B second. As a result, some respondents received the control list (containing three non-sensitive items) first and then the treatment list, while others received a treatment list (containing the sensitive item) first and then the control list. In the end, all respondents answered six different lists (two for each of the child labour measures). The statements used in the three double-list experiments and the design are shown in Table 2.

The effectiveness of the list experiment methodology is based on three assumptions: (i) successful randomisation, (ii) the absence of design effects, and (iii) the absence of ceiling and floor effects. More precisely, individuals allocated to each group must be similar such that on average, they agree with the same number of non-sensitive statements. Second, the addition of the sensitive item must not change the sum of affirmative answers on the control items. Finally, as pointed out by Kuklinski et al., (1997), individuals may provide untruthful answers if they no longer benefit from the privacy of their responses because they either agree or disagree with all the non-sensitive items. We refer to such effects as the ceiling and floor effects, respectively. To eliminate this problem, there should be one non-sensitive item that most participants would agree with and another non-sensitive item that most participants would disagree with. Blair and Imai (2012) also advise choosing non-sensitive items that are related to the topic of the behaviour or opinion investigated in the list experiment to avoid any suspicion on the part of respondents. The choice of the non-sensitive items is key to implementing the list experiment method successfully. In addition, non-sensitive items should be reasonably familiar to the respondent and sufficiently similar in nature and specificity to the sensitive item to avoid introducing bias in the answers (Droitcour et al. 2004) and should not themselves be susceptible to social desirability bias (Hinsley et al. 2019).

We take these items into account in our double-list experiment design. Similarly to the sensitive item, the non-sensitive items on the different lists were selected by ensuring that they related to the sensitive behaviour of interest; we used items related to agriculture and children. A previous survey conducted by the cocoa company was used to select items that would be agreed and disagreed with by everyone on each list. A pilot survey of the 36 farmers confirmed that these items allowed for the absence of floor and ceiling effects. The success of randomisation (assumption (i)) was assessed by comparing a number of individual socio-demographic characteristics between the treated and control groups.

In addition, we implemented two statistical tests to verify whether the addition of the sensitive item modified the answers to the non-sensitive statements (assumption (ii)) (Lépine and Treibich 2020). We ensured that the proportion of individuals in the control group who agreed with no more than y statements ($y=0, 1, 2, 3$) should be greater than this proportion for the treated group and that the latter proportion (for $y=1, 2, 3, 4$) should be greater than the proportion of individuals in the control group who agreed with no more than $y-1$ statements. If this rationale is not the

case, given that individuals in the treatment and control groups are similar on average, it means that individuals in the treatment group modified their answers to the non-sensitive items.

Finally, the potential existence of ceiling and floor effects (assumption (iii)) was investigated by examining the share of individuals in the control group (individuals to whom only three non-sensitive items were presented) for whom $y=0$ or $y=3$.

To estimate the prevalence of sensitive behaviour, we use the following regression:

$$Y_i = \lambda + \beta^l T_i + \varepsilon \quad (1)$$

where Y_i is the number of statements the respondent i agreed with and T_i is a binary variable equal to one if the respondent is assigned to the treatment group and zero otherwise. The average sensitive behaviour prevalence rate is then given by β^l and corresponds to the average difference between the number of statements that the control group and the treatment group agreed with for each list $l=A, B$ separately. When the prevalence obtained from the two lists was not significantly different, we pooled the results from the two lists and estimated:

$$Y_i^l = \alpha + \gamma T_i^l + L_i + \varepsilon \quad (2)$$

where Y_i^l is the number of statements the respondent i agreed with for every two lists l and T_i^l is a binary variable equal to one if the respondent is assigned to the treatment group and zero otherwise in each list. The average sensitive behaviour prevalence rate is then given by γ and corresponds to the average difference between the number of statements that the control and treatment groups agreed with for each list l and L is the list used (A or B). This is estimated using OLS with robust standard errors clustered at the individual level in the linear form and fitted with logistic regression as per Tsai, (2019) in the non-linear and adjusted forms.

In a second step, we estimated how prevalence rates derived from self-reports underestimated the proportion of child labour produced by the list experiment method. To do so, we compared the prevalence estimated from the list method with the prevalence calculated with the direct question. We used a Wald test with the null hypothesis of zero difference.

2.2.2 Colorbox

The colorbox method builds on nonverbal response cards, which were first employed to study sexual behaviour among Ethiopian youth (Lindstrom et al.



Fig. 2 Colorbox coupon

2010), and later, adolescent violence in Burkina Faso (Harling et al. 2021). The colorbox method reduces the risk of confidentiality breaches compared to direct questioning. With the colorbox method, interviewers cannot deduce a respondent's answer to a particular question from their own answers, nor from other questions in the survey, and they never see the coded coupons that respondents use to answer the question.

Respondents are given a coupon (Fig. 2) in a sealed envelope containing two boxes: a white box and a black box. The enumerator informs them that the white box and the corresponding code refer to the answer 'YES', while the black box and the corresponding 6-digit code refer to the answer 'NO'. Respondents are then asked a question about child labour and are instructed to tear off the 6-digit code corresponding to their answer along the dotted line, as shown in Fig. 2. The interviewer enters the 6-digit code (without knowing its correspondence to the box) into the survey software, which determines its validity. There are approximately one million different PIN combinations, only a small proportion of which correspond to YES and NO. As a result, typing errors are unlikely to occur. The envelope given to respondents is sealed by a non-interviewing member of the research team. The codes used vary from respondent to respondent, which prevents enumerators from remembering specific numbers. At the end of the interview, respondents take the coupon with them, ensuring confidentiality between the interviewer and respondent. The envelope contains enough coupons to answer all the questions asked via colorbox. A new coupon is required for each question. Two training questions, where the answer is known, were used to instruct respondents on the methodology for answering questions using this method.

The colorbox method has several advantages over the list experiment. Firstly, the colorbox method allows information to be collected at the individual level, which is empirically simpler than significantly reducing confidence intervals and allows multivariate analysis at the farmer level. It also reduces the length of the survey and is easier to implement in the field, as each question is only one line, as opposed to the double-list experiment, which consists of seven statements plus an explanation of how to answer. Finally, the method allows participants' understanding to be tested using questions for which the correct answer is known.

For both elicitation methods, the mechanism for maintaining participant confidentiality was noted. However, in the case of the colorbox, by explaining the codes to answer the questions, see Appendix 2, the confidentiality mechanism is implied. It was not mentioned that the researchers had the key to the pin codes that allowed them to link responses to farmers, as the data were anonymised before these links were made, and it was not explained which statement from the list experiment was of interest, nor that the researchers had no way of linking responses from the list experiment to individuals. Given the complexity of the colorbox, a practice question was implemented to test whether respondents understood how the method worked, but this was not the case for the list experiment given the simplicity of the question. To capture respondents' perceptions of the method, we asked farmers about (i) their general appreciation of the

method, (ii) their perceptions of the objective of the method, and (iii) their perceptions of the confidentiality of the method and the reason for their response.

3 Results

3.1 Descriptive statistics of the quantitative survey

In the quantitative survey, we surveyed 1741 farmers: 23% in the community of Bianouan, 31% in Affery, 23% in Sinfra and 23% in Blolequin. Households contained seven persons on average, and there were three children aged between 5 and 17 years old in those households; 45.5% of those children were girls, and 78.9% of those children were attending school. On average, 89% of those households had a male head, whose age was 46.9 years old. Forty-four percent of the heads of household never attended school, with an average education level of 4 years (CE1/Year 2). 75% of farmers received activities aiming to prevent child labour. Most farmers (84%) were members of a cooperative as shown in Table 3.

3.2 Validity of the design

The effectiveness of the list experiment methodology is based on three assumptions: (i) the randomisation of the treatment, (ii) the absence of ceiling or floor effects, which would prevent respondents from answering honestly, and (iii) the absence of any design effect, in other words, the fact that adding the sensitive item does not modify the answers regarding the nonsensitive statements. In addition, to pool the results of List A and List B, we needed to prove that the lists are internally consistent. In the next paragraph, we review these hypotheses and check whether they are supported.

Table 3 Descriptive statistics

Variables	Mean	SD	Min	Max
Household size	7.27	3.60	2	30
% head is female	10.79			
Age of head	46.89	12.60	20	93
Years of education of head	4.74	4.94	0	15
Number of children aged 5–17	2.88	1.82	1	15
% children in school	78.94			
% of children who are girls	45.45			
% targeted by child labour prevention activities	75.24			
Wealth	0	1	– 1.25	4.65
% not belonging to a cooperative	15.77			

Wealth is a continuous variable derived from household assets using multi-component analysis ($n = 1741$)

A) Success of randomisation

Appendix 3 shows that randomisation was successful by comparing the means of the outcomes of interest and observable characteristics of those allocated to the list experiment and the colorbox. It also shows the success of randomisation to lists A and B (or treatment and control groups) within the list experiments. The joint significance tests support this.

B) Floor and ceiling effect assumptions and absence of design effect

We also needed to ensure that the list of non-sensitive items provided sufficient privacy for treatment group respondents (assumption (ii)) and that the addition of the sensitive item did not change the responses to the non-sensitive statements (assumption (iii)). In Appendix 4, we estimated the proportion in the control group who disagreed with any statement and answered ‘0’ to the item count list question. As the proportion of people answering ‘0’ in the control group ranged from 0.2 to 4%, there was no question of a floor effect. Regarding the ceiling effect, the proportion of respondents in the control group who answered ‘3’ to the non-sensitive items is extremely low for the lists measuring exposure to agrochemicals and work by children under 13. This proportion is somewhat higher, reaching 6.5% and 15.7% respectively for the two lists measuring the prevalence of children under 18 using sharp tools and carrying heavy loads. These proportions indicate that there may be a

Table 4 Internal consistency of list experiments

Indicator	List	Prevalence	<i>P</i> value
% having children under 18 using sharp tools or carrying heavy loads	1A	11.24 [3.69; 18.80]	0.997
	1B	11.26 [3.05; 19.48]	
% having children under 18 using agrochemicals	2A	16.03 [9.41; 22.66]	0.150
	2B	9.26 [2.81; 15.71]	
% having children under 13 working on plots	3A	10.01 [4.08; 15.94]	0.535
	3B	12.71 [6.59; 18.82]	

95% confidence interval in brackets. We compare the prevalence rates obtained with each list experiment and test the following hypothesis: $\gamma T_i^A = \gamma T_i^B$

privacy issue, i.e., by agreeing to all statements, respondents indicate that they agree with the sensitive statement. The implication is that these lists are not as effective in reducing social desirability bias for all respondents as they could be, meaning that the true proportion is likely to be higher than those found here, although it is likely that only a small proportion would have felt the privacy violation and we do not think the true prevalence would have been much different.

Finally, the difference (row 5) between the proportions of individuals in the treatment group (row 2) and the control group (row 4) agreeing with at least j statements ($j = 1, 2, 3, 4$) is always positive or close to 0, providing evidence for the absence of a design effect (Glynn 2013).

C) Internal consistency of lists A and B

Based on the results presented in Table 4, we found that the two lists used for each indicator of child labour provide prevalence estimates that are not significantly different from each other. These tests provide evidence in support of the internal consistency of the list experiment method and allow us to pool the results to analyse results using the double list experimental design in the latter section.

D) Understanding and perceived confidentiality

The two methods were perceived as highly confidential: 95.64% of the farmers stated that the colorbox was confidential and 96.64% of the farmers stated that the list experiment was confidential. In addition, only ten farmers (1.8%) were unable to answer the colorbox test questions correctly. These observations were excluded from the rest of the analysis.

Table 5 Child labour prevalence

Child labour measures	Self-report	List experiment	Colorbox
During this campaign, children under 18 use sharp tools such as machetes or carry heavy loads like a 10 L water tank on your plots?	8.88 [7.47; 10.14] <i>N</i> = 1737	11.26 [6.27; 16.24] <i>N</i> = 2380 (farmers = 1190)	10.54 [7.94; 13.13] <i>N</i> = 541
During this campaign, children under 18 are present when agrochemical are used on your plots?	3.17 [2.34; 3.99] <i>N</i> = 1737	12.65 [8.44; 16.86] <i>N</i> = 2380 (farmers = 1190)	6.48 [4.40; 8.56] <i>N</i> = 541
During this campaign, children under 13 work on your plots?	4.97 [3.94; 5.98] <i>N</i> = 1732	11.36 [7.28; 15.43] <i>N</i> = 2380 (farmers = 1190)	9.98 [7.45; 12.52] <i>N</i> = 541

Means are reported with their 95% confidence intervals in brackets. For the list experiment, we pooled the two lists obtained from the 1190 participants randomised to this method. We removed the 10 respondents who did not understand the colorbox method to estimate the prevalence of child labour using this method. *N* stands for the number of observations. Appendix 6 contains non-linear and adjusted models

3.3 Estimation of child labour prevalence

Table 5 presents the estimated child labour prevalence using the direct question and the indirect elicitation methods. We found that the indirect methods lead to a greater prevalence of child labour in comparison to the direct question and that these prevalence estimates are significantly different for two out of three of these indicators. However, the degree of misreporting differed depending on the child labour prevalence estimated.

Mean child labour prevalence estimates were higher with the list experiment than with colorbox, although the difference was not statistically significant. This is in line with the fact that the list experiment does provide the greatest degree of confidentiality for respondents. The colorbox reduced the reporting difference between the mean prevalence elicited via the direct question and the list experiment by between 35% and 78% (Appendix 5). Despite the list experiment using the double list design and two-thirds of the sample, the colorbox estimates had a lower variance using only a third of the sample.

For two of the three indicators, there was no difference in the prevalence of child labour obtained using colorbox and the list experiment; the only difference was that the presence of children in the use of agrochemicals was twice as high using the list experiment method as using colorbox.

In Appendix 6, we repeat the list of experiment data using non-linear estimation and adjust for characteristics. These findings support the unadjusted figures reported in Table 3, with non-linear estimates resulting in a slightly higher estimated prevalence (Question 1 = 12.9%, Question 2 = 14.3%, Question 3 = 12.9%). Adjustment for farmer characteristics supports the results for questions 1 and 3, but question 2 on agrochemicals gives a prevalence that is not statistically different from 0 after adjustment.

3.4 Qualitative reasons for misreporting child labour

Qualitative findings surrounding truthfulness of child labour reporting are found in Appendix 7. They highlight that farmers are unwilling to report child labour to all stakeholders but primarily to cocoa companies and cooperatives due to perceived sanctions.

We further investigated the reasons for which farmers were reluctant to reveal child labour in the qualitative survey. Three main themes emerged surrounding the misreporting of child labour: (i) the high perceived sanctions from the government, the community and the cooperative, (ii) the good knowledge of the risks associated with child labour on children's health outcomes and (iii) the trustworthiness of the stakeholders asking the questions.

3.4.1 Perception of sanctions

Most persons interviewed knew about the fact that child labour was illegal. The information was obtained through TV, radio and local awareness campaigns.

Sinfra ID 11: *“They had signs in Sinfra that said not to use children, so we know there is a law.”*

Most farmers perceived these sanctions to be enforced. The main fear of farmers was to be arrested and to go to jail, and many of them reported knowing people who went to jail. Farmers also reported that police officers would make unplanned visits to the fields.

Sinfra ID 4: *“They said that in Soubré there was a man who had taken the children who had formed a “society” (a sort of cooperative) and they went to work in his field, and they took the man to the police station. He was put in prison.”*

Bianouan ID 8: *“Last year, there was a bus that left Abidjan and came, and the policemen went into the fields, so we do not make our children work.”*

In addition, there was a low level of trust towards the police and the legal system. Farmers mentioned that the police could arrest them and put them in jail without going to court.

Bolequin ID 6: *“It’s a problem because the police has the right to check everywhere, if they see they can imprison you directly, without asking any question.”*

Outside penal sanctions, farmers reported that they would face sanctions from the village and the community if they worked with children:

Sinfra ID 5: *“Yes, here, when there is this kind of problem, the village chief says, you are being fined, that you are going to pay this or that, that you have to pay this or that to the village. You can be fined CFAF 5,000 to 10,000 when the sanction is serious.”*

They also reported some sanctions by the cooperative, mainly the loss of certification and the impossibility of selling cocoa.

Bianouan ID 9: *“In the cooperative, when you refuse, your cocoa is not taken, you go and deliver elsewhere because the cooperative has its rules.”*

3.4.2 Knowledge of risks associated with child labour

Farmers highlighted many physical and mental health risks associated with child labour, and for those who still did it, the reasons were financial (saving on paid labourers ‘manoeuvre’) and training their children for their future work as cocoa farmers. Most farmers agreed that there were many costs associated with child labour, such as health expenses due to an accident, the opportunity cost of not having a child in school, and the fact that children could not be used for all types of farm work.

Most parents were aware of many risks to children, such as snake and insect bites, cuts, falling wood, drowning, burns, fatigue and illness:

Sinfra ID 13: *“Let’s say with themachete, the child can get hurt, the daba also the same, because it is sharp. Using these tools can hurt the child. When children go to the field, you are not near them, there are many things that can happen to them, like snakes, scorpions, hairy spiders, all that, they are dangerous animals. Then, by using the child to work, his body can become deformed in a way. The pain can act on him.”*

Another also reported mental health issues because of forced or difficult labour:

Bolequin ID 6: *“It can have diseases and then the child’s brain can be messed up for it.”*

Sinfra ID 5: *“When you make a child do work that is beyond him, it delays his progress, his memory is threatened, when you make him do work that is beyond him, when you see him coming, he is already afraid; you put fear in him.”*

Despite this, the main reasons for child labour were to train children to work in cocoa production and to save the costs of daily workers:

Bianouan ID 6: *“They do it in the permanence of the work so that tomorrow the children also accept to work in the plantations, so that when they grow up, they continue the cocoa production.”*

Bianouan ID 6: *“The majority of those who do this are doing it to avoid taking daily workers to work, since they know that if they take a labourer they will pay for it, but if they take their child they will not pay.”*

3.4.3 Trustworthiness

Farmers were unwilling to share with other farmers because of a low level of trust in them due to perceived rivalry and jealousy:

Bolequin ID 2: *“Those who don’t have children, if I have children working, the other farmer will try to report to the cooperative or to other person, make criticisms to say that I force the children to work in the field. Because of jealousy, since he doesn’t have any, he will go and criticise him up there and then they will come and create problems for me.”*

However, many believed that there was less risk in revealing child labour to researchers who were considered external and trustworthy.

Bolequin ID 3: *“No that’s not a problem because academic, everyone knows that he is not an authority, he’s just here advise.”*

Bolequin ID 5: *“It’s not a problem because the academic is looking for his degree that’s why, he came to talk with you.”*

Bolequin ID 6: *“Yes, because there are some planters, you go to see him, he has children from 5 to 13 years old, he is afraid to say, it can scare because after his diploma, if he has had this kind of work, he can come to him and then you who made your child work there you have a problem.”*

But farmers did not want enumerators to be able to see if children were working on the plots and if their physical presence in the community was disturbing.

Bolequin ID 5: *“Since you are here (enumerator), to make it good, I have to arrange the work so that it’s clean, so that we don’t see child labour in the fields.”*

Qualitative findings highlighted the importance of using indirect elicitation methods since direct self-report will lead to underreporting. At this stage, it was not clear if the colorbox method could solve social desirability bias, given that participants had a mixed level of trust toward researchers. For this reason, we both tested the list experiment and colorbox methods.

The results also indicated the potential for bias in surveys conducted by different stakeholders. Those figures collected by cocoa companies or cooperatives may be more susceptible to sensitivity bias than our findings as academic researchers. This highlights the necessity to collect quantitative data using a team of independent fieldworkers instead of using cooperative or cocoa company staff. Conversely, social

desirability bias may be influencing the qualitative responses, with farmers indicating that academic researchers are more trustworthy because academic researchers were asking the questions.

3.5 Determinants of child labour

The determinants of child labour were then analysed using direct questions, colorbox and list experiments, methodologies discussed in Imai (2010) and Tsai (2019) and as demonstrated in quasi-experimental circumstances Cust et al., (2023). A direct question revealed that farmers who had engaged in child labour prevention activities were less likely to report using children. This reduction of 3–6.9 percentage points was statistically significant at 5% for the second and third measures of child labour (Table 6). However, this was no longer the case when using the indirectly elicited indicators. While some determinants were statistically significant using direct questions, such as the zone, belonging to a cooperative, and the number of children, they were no longer statistically significant using the colorbox and list experiment methods. In fact, no important determinant of child labour was consistently identified using either the colorbox or list experiment methods.

The lack of difference between the colorbox and list experiment compared to direct questioning of those targeted by child labour prevention activities suggests that they may only be effective in educating farmers about the consequences of child labour, increasing their likelihood of not reporting it, but not actually reducing the level of child labour.

4 Discussion

Qualitative evidence suggests that this misreporting is due to fear of sanctions by farmers who are aware of the illegality of child labour and believe that sanctions are severe and effectively enforced. This point also points to the use of indirect survey methods, but they are unlikely to have eliminated all sources of sensitivity bias. Regionally representative evidence from NORC finds much higher levels of child labour at the child level (31% using sharp tools, 26% carrying heavy loads and 19% exposed to agrochemicals), suggesting that our quantitative findings at the household level still underestimate the true prevalence (Sadhu et al. 2020). Qualitative findings highlighting parents' good knowledge of the consequences of child labour on their mental and physical health, suggesting a negative self-image, leading to a greater social desirability bias in our study compared to NORCs. Our study also found that when comparing the determinants of child labour using direct versus indirect questioning, there was some evidence that misreporting was greater among farmers targeted by a child labour reduction intervention when questioned directly, but this effect disappeared when social desirability bias was reduced using the colour box. This suggests that child labour interventions may improve understanding of the consequences and therefore encourage farmers not to report, but do not actually reduce child labour.

Table 6 Determinants of child labour

	Direct questions			Colorbox questions			List experiment		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Reference zone: Bianouan									
Aifery	-0.101*** (0.020)	-0.040*** (0.013)	-0.087*** (0.016)	0.004 (0.039)	0.014 (0.029)	-0.043 (0.038)	0.054 (0.081)	-0.106 (0.070)	0.033 (0.059)
Sinfra	0.020 (0.026)	-0.003 (0.015)	-0.000 (0.022)	0.053 (0.047)	0.013 (0.034)	0.026 (0.044)	0.127 (0.092)	0.009 (0.067)	0.145** (0.067)
Biolequin	-0.119*** (0.020)	-0.046*** (0.012)	-0.089*** (0.017)	0.001 (0.039)	0.010 (0.035)	0.000 (0.041)	0.052 (0.091)	0.113 (0.076)	0.012 (0.066)
Not belonging to a cooperative	0.086*** (0.024)	0.044*** (0.016)	0.004 (0.016)	0.062 (0.048)	0.041 (0.035)	0.041 (0.042)	0.154*** (0.078)	-0.048 (0.069)	0.064 (0.063)
Received child labour prevention activities	-0.069*** (0.019)	-0.030** (0.013)	-0.046*** (0.016)	-0.038 (0.039)	0.013 (0.030)	-0.034 (0.038)	0.053 (0.070)	0.014 (0.061)	0.023 (0.055)
Household size	-0.004 (0.002)	0.000 (0.002)	-0.003 (0.002)	0.001 (0.005)	-0.002 (0.004)	0.002 (0.005)	0.014 (0.010)	-0.006 (0.008)	0.003 (0.007)
Head is female	0.016 (0.024)	0.034* (0.018)	0.034* (0.021)	-0.013 (0.041)	0.013 (0.038)	-0.016 (0.032)	0.108 (0.079)	0.004 (0.068)	-0.058 (0.068)
Age of head	0.001 (0.001)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.002)	0.002 (0.002)	-0.004** (0.002)
Years of education of head	-0.002* (0.001)	-0.000 (0.001)	-0.001 (0.001)	-0.005* (0.003)	-0.005** (0.002)	-0.005* (0.003)	0.005 (0.007)	0.010** (0.005)	-0.006 (0.005)
Number of children aged 5-17	0.012** (0.005)	0.002 (0.004)	0.012** (0.005)	-0.008 (0.009)	0.011 (0.009)	0.009 (0.009)	-0.027 (0.018)	0.003 (0.015)	0.004 (0.014)
% of children who are girls	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.001 (0.000)	-0.001 (0.000)	-0.000 (0.000)	-0.000 (0.001)	0.000 (0.001)	-0.001 (0.001)

Table 6 (continued)

	Direct questions			Colorbox questions			List experiment		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
% of children in school	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.001)	-0.000 (0.000)	-0.000 (0.000)	-0.002*** (0.001)	0.002*** (0.001)	-0.001 (0.001)
Wealth	0.007 (0.008)	0.002 (0.004)	0.001 (0.005)	-0.001 (0.013)	0.014 (0.012)	0.020 (0.013)	0.037 (0.035)	-0.014 (0.028)	-0.004 (0.025)
Constant	0.138*** (0.038)	0.078*** (0.025)	0.141*** (0.032)	0.152* (0.079)	0.144** (0.066)	0.190** (0.079)	0.152 (0.161)	-0.170 (0.140)	0.330*** (0.120)
Observations	1,735	1,735	1,730	551	550	551	2,376	2,376	2,376
R-squared	0.079	0.037	0.062	0.035	0.029	0.037			

Indicators were denoted as follows: (1) During this campaign, are there children under 18 who use sharp tools such as machetes or carry heavy loads like a 10 L water tank on your plots? (2) During this campaign, are there children under 18 who are present when agrochemicals are used on your plots? (3) During this campaign, are there children under 13 who work on your plots? Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. R-squared not reported using linear models (Tsai 2019)

It is possible that the order of the questions means that there is a potential priming of respondents as to which questions are coming next, which could bias their response. To minimise this effect, we did not tell respondents that the questions were about child labour, then started with the indirect methods, the list experiment or the colorbox method, and finally the direct questions, all in the same order for participants assigned to the colorbox or list experiment arm. The three types of child labour questions were also asked in the same order within each method. Given that respondents found the methods to be highly protective of identity, we do not believe that the bias from the order of the questions is large. There is a possibility that once the surveys began, there was a local spillover of the content of the questionnaire. However, each village was surveyed in a single day by large enumeration teams, limiting the opportunity for communication and warning between farmers. Spillover of primed knowledge about child labour between villages is also a possibility. We found no evidence that misreporting changed over time, suggesting that there was no spillover of knowledge ahead of the survey teams, but that even if there was, it did not affect the prevalence measures (results available on request).

Our results rely on the good design of the indirect elicitation methods. We show that the double list experimental design led to internally consistent results, which reinforces the confidence in the results obtained by this method, which is important given that the prevalence of child labour we obtained in this study was lower than that obtained by Jouvin (2023). However, Jouvin (2023) was measuring child labour in broader terms ('help harvesting and breaking pods', 'preparing the farm' and 'any help' from those under 16). While colorbox led to a lower prevalence of our second estimator compared to the list experiment, our study showed the promising use of the colorbox method to elicit sensitive questions since it allowed us to uncover a large proportion of the misreporting obtained, using the list experiment method as a benchmark. In addition, colorbox allows us to minimise statistical noise by allowing researchers to obtain information regarding sensitive questions at the individual level. There was evidence that farmers perceived this method to be highly confidential. A key contribution of this paper is the feasibility of using the colorbox and we recommend this method over the list experiment in future research investigating the effect of interventions to reduce child labour.

Although our study found a different prevalence of child labour than previously reported, it is important to note that all studies use different definitions and measures of child labour. Our study focused on hazardous activities, which may explain the lower prevalence found compared to other studies (Jouvin 2023; Lichand and Wolf 2022). Our study provides a new prevalence of child labour estimated using an internally consistent list experiment design and a newly developed non-verbal method; however, it is important to note that it has some limitations. First, it would have been preferable to also randomly allocate the direct self-report question to avoid possible response bias since the direct questions were asked after the indirect methods. This was not feasible given our sample size. Future studies should investigate whether the order of the direct and indirect methods affects the prevalence obtained with the second method. Second, the indirect methods used do not allow the intensity of child labour (e.g., number of hours worked) to be captured, as they are designed to measure binary

outcomes. Third, the use of the double list experiment design may induce fatigue from respondents. Consequently, researchers should conduct formative studies to ensure that the survey design is feasible and well accepted by survey participants. A final limitation is the external validity, particularly of the child labour prevalence estimates, which should not be applied to areas outside the regions studied. As the study was primarily a methodological study of the use of indirect survey methods, the sample was selected to provide a broad range of types of cocoa farmers in Côte d'Ivoire, rather than a representative sample of the farmer population. We are confident that our methodological conclusions regarding the determinants of bias and the performance of the measures are applicable to other geographical areas.

Given the difficulties in obtaining accurate responses, numerous recent studies have employed direct questioning of children to ascertain the prevalence of child labour (Dillon 2010). A study by NORC on the prevalence of child labour in selected cocoa-growing countries employed a number of methodologies to ascertain the prevalence of child labour at the child level. These included direct questioning of children residing in households engaged in cocoa farming, as well as the use of household head and school surveys (Sadhu et al. 2020). More recently, Lichand and Wolf (2022) also combined data collected from children based on independent reports from primary school children randomly selected from a sample of schools and their parents in two cocoa-producing regions of Côte d'Ivoire. They found a higher prevalence of child labour when obtained from children (45%) than from parents (16%). However, determining prevalence from information collected from children and adults is difficult to compare due to differences in definitions and denominators. For instance, the actual prevalence of establishments employing child labour and the number of children may be significantly disparate. Furthermore, surveying children raises significant ethical concerns (Devries et al. 2015). The process of surveying minors requires both the assent of the minors themselves and the consent of their parents or guardians. Obtaining parental consent may, however, place children at risk of violence, particularly in countries where violence against children is common (Devries et al. 2021; Hillis et al. 2016). Previous studies have not addressed these ethical challenges. Furthermore, while school-based surveys have the advantage of ensuring that parents and children are interviewed independently, they require administrative approval and may be prone to serious sample selection problems. If children working in the cocoa fields are also out of school, obtaining child labour levels from school surveys may also lead to underreporting of child labour. In addition to the problems mentioned above, surveying children can introduce other biases. For example, children may have lower cognitive development than adults (Borgers et al. 2000). Furthermore, children may be susceptible to social desirability bias if they are aware of the illegality of child labour and the risks to their parents. Conversely, they may over-report if they expect to receive benefits or assistance as a result of disclosing their child labour. This is of particular relevance in Côte d'Ivoire, where some child labour prevention strategies are specifically designed to benefit children, for instance, by providing school supplies and waiving school fees.

The results of this study have several implications for future research on child labour. First, it was found that some aspects of child labour were more sensitive to farmers than others. In particular, the presence of children when agrochemicals are used was more underreported when direct questions were used than the use of sharp tools or the carriage of heavy loads. This suggests that the choice of child labour indicator should be carefully made if elicited directly. Further research is needed to capture the perception of cocoa farmers regarding different aspects of child labour. The observed bias following targeted activities to reduce child labour suggests that future evaluations of the effectiveness of interventions to reduce child labour should avoid direct questions to elicit child labour prevalence. This is because such questions may lead to biased estimates that capture only changes in reporting and not underlying child labour. Consequently, it is important to use indirect elicitation methods in impact evaluations. Finally, we urge researchers engaged in the study of child labour to employ indirect elicitation methods, to rely on independent research teams and to refrain from utilising the personnel of the cooperative in the recruitment of research participants. Indeed, the qualitative study revealed that farmers would be disinclined to disclose instances of child labour if data were collected by the cooperative or by cocoa companies, given that they perceived these parties to be working with the government and to be highly engaged in the fight against child labour.

Policy that could be effective in preventing child labour should focus on raising the marginal productivity of workers and increasing the opportunity cost of sending children to work. It is notable that a significant proportion of cocoa production in Côte d'Ivoire is without mechanisation and very labour-intensive with low productivity per worker. This leads to head farmers employing daily labourers for very low wages, which creates a strong temptation to employ children. By investing in more efficient farm techniques that rely less on human labour, the benefit of employing children over adults will diminish. Second, in the short term, raising the opportunity cost of work for children through incentivising school attendance, such as conditional cash transfer programmes (public school is already free in Côte d'Ivoire), could remove many children from farms should the transfer be greater than the average adult wage on the farm.

In this study, three child labour indicators were elicited using two indirect elicitation methods. The results indicated that indirect methods led to a significantly larger prevalence, up to four times larger. Our findings suggest that questioning adults using indirect elicitation methods might still yield underestimates in child labour prevalence compared to questioning children directly. This is because indirect elicitation does not eliminate all sources of sensitivity bias. This is supported by qualitative research results showing that farmers were terrified of sanctions and had a low level of trust. Qualitative research elucidated the underlying causes of this misreporting and revealed that farmers were aware of the legal sanctions in place and considered them to be enforced. The most severe sanctions they described included a jail sentence, loss of certification and the impossibility of selling their agricultural output. Given the importance of tracking child labour indicators and evaluating the impact

of interventions reducing child labour, further work should be conducted with farmers to foster trust and ensure safe collaborations. In the meantime, while child labour prevalence could be better enumerated when asked of children, further research is required to evaluate the ethical risks associated with scaling up child surveys.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00148-024-01054-3>.

Acknowledgements We acknowledge the farmers who participated in this study and the leaders of the cooperative for their involvement in the study. We also thank editor Kompal Sinha and three reviewers for helpful comments.

Declarations

Conflict of interest The project was designed and carried out in collaboration with Barry Callebaut and the International Cocoa Initiative. It was funded by The Netherlands Enterprise Agency (RVO) and Barry Callebaut.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agüero JM, Frisancho V (2022) Measuring violence against women with experimental methods. *Econ Dev Cult Change* 70:1565–1590. <https://doi.org/10.1086/714008>
- Asadullah MN, De Cao E, Khatoun FZ, Siddique Z (2021) Measuring gender attitudes using list experiments. *J Popul Econ* 34:367–400. <https://doi.org/10.1007/s00148-020-00805-2>
- Beegle K, Carletto C, Himelein K (2012) Reliability of recall in agricultural data. *Journal of Development Economics, Symposium on Measurement and Survey Design* 98:34–41. <https://doi.org/10.1016/j.jdeveco.2011.09.005>
- Bell SO, Bishai D (2019) Can a list experiment improve validity of abortion measurement? *Stud Fam Plann* 50:43–61. <https://doi.org/10.1111/sifp.12082>
- Blair G, Imai K (2012) Statistical Analysis of List Experiments. *Polit Anal* 20:47–77. <https://doi.org/10.1093/PAN/MPR048>
- Borgers, N., de Leeuw, E., Hox, J., 2000. Children as respondents in survey research: cognitive development and response quality. *BMS: Bulletin of Sociological Methodology / Bulletin de Méthodologie Sociologique* 60–75.
- Castilla C, Murphy DMA (2023) Bidirectional intimate partner violence: evidence from a list experiment in Kenya. *Health Econ* 32:175–193. <https://doi.org/10.1002/hec.4616>
- Chingandu, N., Kouakou, K., Romain, A., Gutierrez, O., Brown, J., 2017. Unexpected genome variability at multiple loci suggests cacao swollen shoot virus comprises multiple, divergent molecular variants. *Journal of Emerging Diseases and Virology* 3. <https://doi.org/10.16966/2473-1846.128>
- Chuang, E., Dupas, P., Huillery, E., Seban, J., 2021. Sex, lies, and measurement: consistency tests for indirect response survey methods. *Journal of Development Economics* 148. <https://doi.org/10.1016/j.jdeveco.2020.102582>
- Côte d'Ivoire United States Department of Labor (DOL), 2015.

- Cullen C (2023) Method matters: the underreporting of intimate partner violence. *The World Bank Economic Review* 37:49–73. <https://doi.org/10.1093/wber/lhac022>
- Cust H, Lépine A, Treibich C, Powell-Jackson T, Radice R, Tidiane Ndour C (2023) Trading HIV for sheep: risky sexual behavior and the response of female sex workers to Tabaski in Senegal. *Health Economics* n/a. <https://doi.org/10.1002/hec.4756>
- Devries KM, Child JC, Elbourne D, Naker D, Heise L (2015) “I never expected that it would happen, coming to ask me such questions”: ethical aspects of asking children about violence in resource poor settings. *Trials* 16:516. <https://doi.org/10.1186/s13063-015-1004-7>
- Devries K, Balliet M, Thornhill K, Knight L, N’djoré, Y.A.B., N’guessan, D.G.F., Merrill, K.G., Dally, M., Allen, E., Hossain, M., (2021) Can the ‘Learn in peace, educate without violence’ intervention in Côte d’Ivoire reduce teacher violence? Development of a theory of change and formative evaluation results. *BMJ Open* 11:e044645
- Dillon, A., 2010. Measuring child labor: comparisons between hours data and subjective measures. *Research in Labor Economics* 135–159.
- Dillon A, Bardasi E, Beegle K, Serneels P (2012) Explaining variation in child labor statistics. *Journal of Development Economics, Symposium on Measurement and Survey Design* 98:136–147. <https://doi.org/10.1016/j.jdeveco.2011.06.002>
- Droitcour, J., Caspar, R.A., Hubbard, M.L., Parsley, T.L., Visscher, W., Ezzati, T.M., 2004. The item count technique as a method of indirect questioning: a review of its development and a case study application, in: Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (Eds.), *Wiley Series in Probability and Statistics*. Wiley, pp. 185–210. <https://doi.org/10.1002/9781118150382.ch11>
- Gilligan, D.O., Hidrobo, M., Leight, J., Tambet, H., 2024. Using a list experiment to measure intimate partner violence: cautionary evidence from Ethiopia. *Applied Economics Letters* 0, 1–7. <https://doi.org/10.1080/13504851.2024.2308579>
- Glynn AN (2013) What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opin Q* 77:159–172
- Gonzalez-Ocantos E, de Jonge CK, Meléndez C, Osorio J, Nickerson DW (2012) Vote buying and social desirability bias: experimental evidence from Nicaragua. *American Journal of Political Science* 56:202–217. <https://doi.org/10.1111/j.1540-5907.2011.00540.x>
- Government of Côte d’Ivoire, Government of Ghana, U.S. Department of Labor, International Chocolate and Cocoa Industry, 2021. CLCCG REPORT: 2010–2020 Efforts to Reduce Child Labor in Cocoa.
- Guarcello L, Kovrova I, Lyon S, Manacorda M, Rosati FC (2010). Towards Consistency in Child Labour Measurement: Assessing the Comparability of Estimates Generated by Different Survey Instruments. <https://doi.org/10.2139/ssrn.1777103>
- Haber, N., Harling, G., Cohen, J., Mutevedzi, T., Tanser, F., Gareta, D., Herbst, K., Pillay, D., Bärnighausen, T., Fink, G., 2018. List randomization for eliciting HIV status and sexual behaviors in rural KwaZulu-Natal, South Africa: a randomized experiment using known true values for validation. *BMC Medical Research Methodology* 2018 18:1 18, 1–12. <https://doi.org/10.1186/S12874-018-0507-9>
- Hadji M, Asghari F, Yunesian M, Kabiri P, Fotouhi A (2016) Assessing the prevalence of publication misconduct among Iranian authors using a double list experiment. *Iran J Public Health* 45:897–904
- Harling G, Bountogo M, Sié A, Bärnighausen T, Lindstrom DP (2021) Nonverbal response cards reduce socially desirable reporting of violence among adolescents in rural Burkina Faso: a randomized controlled trial. *J Adolesc Health* 68:914–921
- Hillis, S., Mercy, J., Amobi, A., Kress, H., 2016. Global prevalence of past-year violence against children: a systematic review and minimum estimates. *Pediatrics* 137.
- Hinsley A, Keane A, St. John, F.A.V., Ibbett, H., Nuno, A., (2019) Asking sensitive questions using the unmatched count technique: applications and guidelines for conservation. *Methods Ecol Evol* 10:308–319. <https://doi.org/10.1111/2041-210X.13137>
- Holbrook AL, Krosnick JA (2010) Social desirability bias in voter turnout reports: tests using the item count technique. *Public Opin Q* 74:37–67. <https://doi.org/10.1093/poq/nfp065>
- Ibrahim A, Abdalla SM, Jafer M, Abdelgadir J, de Vries N (2019) Child labor and health: a systematic literature review of the impacts of child labor on child’s health in low- and middle-income countries. *J Public Health* 41:18–26. <https://doi.org/10.1093/pubmed/fdy018>
- Imai K (2011) Multivariate regression analysis for the item count technique [WWW Document]. *J Am Stat Assoc*. <https://doi.org/10.1198/jasa.2011.ap10415>
- Jouvin M (2023) Addressing social desirability bias when measuring child labor use: an application to cocoa farms in Côte d’Ivoire. *World Bank Econ Rev*. <https://doi.org/10.1093/wber/lhad030>

- Karlan DS, Zinman J (2012) List randomization for sensitive behavior: an application for measuring use of loan proceeds. *J Dev Econ* 98:71–75. <https://doi.org/10.1016/j.jdeveco.2011.08.006>
- Krumpal I (2013) Determinants of social desirability bias in sensitive surveys: a literature review. *Qual Quant* 47:2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>
- Kuklinski JH, Cobb MD, Gilens M (1997) Racial attitudes and the “New South.” *The Journal of Politics* 59:323–349. <https://doi.org/10.1017/S0022381600053470>
- LaBrie JW, Earleywine M (2000) Sexual risk behaviors and alcohol: higher base rates revealed using the unmatched-count technique. *J Sex Res* 37:321–326. <https://doi.org/10.1080/00224490009552054>
- Lax JR, Phillips JH, Stollwerk AF (2016) Are survey respondents lying about their support for same-sex marriage? Lessons from a list experiment. *Public Opin Q* 80:510–533. <https://doi.org/10.1093/poq/nfv056>
- Lépine A, Treibich C (2020) Risk aversion and HIV/AIDS: evidence from Senegalese female sex workers. *Soc Sci Med* 256:113020. <https://doi.org/10.1016/j.socscimed.2020.113020>
- Lépine A, Treibich C, D’Exelle B (2020) Nothing but the truth: consistency and efficiency of the list experiment method for the measurement of sensitive health behaviours. *Soc Sci Med* 266:113326. <https://doi.org/10.1016/j.socscimed.2020.113326>
- Lépine A., Toh, W.Q., Treibich, C., 2024. Colorbox: a novel method for eliciting sensitive behaviours in face-to-face interviewer-led surveys. Working Paper.
- Lichand, G., Wolf, S., 2022. Measuring child labor: whom should be asked, and why it matters.
- Lindstrom DP, Belachew T, Hadley C, Hattori MK, Hogan D, Tessema F (2010) Nonmarital sex and condom knowledge among Ethiopian young people: improved estimates using a nonverbal response card. *Stud Fam Plann* 41:251–262. <https://doi.org/10.1111/j.1728-4465.2010.00251.x>
- McKenzie D, Siegel M (2013) Eliciting illegal migration rates through list randomization. *Migration Studies* 1:276–291. <https://doi.org/10.1093/migration/mnt018>
- Moseson, H., Jayaweera, R., Huber-Krum, S., Garver, S., Norris, A., Gerdt, C., 2021. Reducing under-reporting of abortion in surveys: results from two test applications of the list experiment method in Malawi and Senegal. *PLoS ONE* 16. <https://doi.org/10.1371/journal.pone.0247201>
- Nederhof AJ (1985) Methods of coping with social desirability bias: a review. *Euro J Social Psych* 15:263–280. <https://doi.org/10.1002/ejsp.2420150303>
- Our World in Data., 2021. Cocoa bean production.
- Sadhu, S., Kysia, K., Onyango, L., Zinnes, C., Lord, S., Monnard, A., Arellano, I.R., 2020. NORC final report: assessing progress in reducing child labor in cocoa production in cocoa growing areas of Côte d’Ivoire and Ghana. NORC at the University of Chicago.
- Treibich C, Lépine A (2019) Estimating misreporting in condom use and its determinants among sex workers: evidence from the list randomisation method. *Health Econ* 28:144–160. <https://doi.org/10.1002/hec.3835>
- Tsai C (2019) Statistical analysis of the item-count technique using Stata. *Stand Genomic Sci* 19:390–434. <https://doi.org/10.1177/1536867X19854018>
- Valente C, Toh WQ, Jalingo I, Lépine A, de Paula Á, Miller G (2024) Are self-reported fertility preferences biased? Evidence from indirect elicitation methods. *Proc Natl Acad Sci* 121(34):e2407629121

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Aurelia Lepine¹ · Ariane Ndiore² · Carole Treibich³ · Henry Cust¹ · Laurent Foubert⁴ · Megan Passey⁴ · Selina Binder⁵

✉ Aurelia Lepine
a.lepine@ucl.ac.uk

Ariane Ndiore
arianendjore@gmail.com

Carole Treibich
carole.treibich@univ-grenoble-alpes.fr

Henry Cust
h.cust@ucl.ac.uk

Laurent Foubert
l.foubert@cocoainitiative.org

Megan Passey
m.passey@cocoainitiative.org

Selina Binder
selina_binder@barry-callebaut.com

¹ Institute for Global Health, University College London, London, UK

² Université Félix Houphouët-Boigny, Abidjan, Cote d'Ivoire, Switzerland

³ Univ. Grenoble Alpes, CNRS, INRAE, Grenoble INP, GAEL, Grenoble 38000, France

⁴ International Cocoa Initiative, Geneva, Switzerland

⁵ Barry Callebaut, Zurich, Switzerland